

OPTIMAL SPACE-TIME COVERAGE AND EXPLORATION COSTS IN GROUNDWATER MONITORING NETWORKS

L. M. NUNES^{1*}, M. C. CUNHA² and L. RIBEIRO³

¹ Faculty of Marine and Environmental Sciences, University of Algarve, Campus de Gambelas, Faro, Portugal; ² Civil Engineer Department, University of Coimbra, Pinhal de Marrocos, Coimbra, Portugal; ³ Instituto Superior Técnico, Lisbon Technical University, Av. Rovisco Pais, Lisboa Codex, Portugal

(* author for correspondence, e-mail: lnunes@ualg.pt)

(Received 6 June 2002; accepted 1 April 2003)

Abstract. A method to determine the optimal subset of stations from a reference level groundwater monitoring network is proposed. The method considers the redundancy of data from historical time series, the times associated with the total distance required to run through the entire monitoring network, and the sum of the times for each monitoring station. The method was applied to a hypothetical case-study consisting of a monitoring network with 32 stations. Cost-benefit analysis was performed to determine the number of stations to include in the new design versus loss of information. This optimisation problem was solved with simulated annealing. Results showed that the relative reduction in exploration costs more than compensates for the relative loss in data representativeness.

Keywords: groundwater, monitoring, network, simulated annealing

1. Introduction

The problem of reducing the dimension of an existing groundwater monitoring network (GMN) is addressed in this article. Both the quantity and quality of the data and the exploration costs are included in the objective function. Data quality includes the representativeness of the parameter spatial field and temporal variation. The parameter spatial field is considered to be representative for the phenomenon under study if, with the amount of information available, the spatial variation is *best* reproduced with a well-defined classification method. Three methods have been proposed to obtain optimal GMN networks: (i) variance reduction; (ii) coupled simulation-statistical analysis; (iii) information transmission (transinformation). The variance reduction (VR) methods were introduced in early works by Delhomme (1978), who used the so-called fictitious point method (generally used to assess the quality of covariance models when estimating with kriging), to determine the optimal location of rain gauges.

VR methods use the variance of the estimation error (σ_E^2) as an indicator of the accuracy of the estimated values. In geostatistics, the mathematical definition of σ_E^2 means that its value does not depend on the actual values of the measured variables, but rather on the relative spatial distribution of the measuring locations. Therefore,



Environmental Monitoring and Assessment **93**: 103–124, 2004.

© 2004 Kluwer Academic Publishers. Printed in the Netherlands.

one may use σ_E^2 to indicate the spatial distribution optimality of a sampling network by testing all the combinations of available sampling locations, and choosing the combination that minimises σ_E^2 . Some additional assumptions must be made about the probability-density function for the estimation error: the estimation errors at any location in space are normally distributed – this may not always be the case (Journel, 1987) – with zero mean and a standard deviation equal to the square root of the estimation variance. A VR method is applied in this article. A short review of the other two other optimisation methods is given below.

The coupled simulation-statistical analysis includes a wide range of techniques that have arisen from the need to optimise the spatial distribution of containment or pump-and-treatment wells in groundwater pollution problems. Meyer and Downey (1988) proposed a method for determining the best location for monitoring wells, following the works of Massmann and Freeze (1987a, b) in a risk-cost-benefit analysis for waste management facilities. The method's intention was to select the networks that maximize the probability of detection in the face of uncertainty. However, its practical applicability is hindered by the extreme simplicity of the analytical model used by Meyer and Downey (1988). Since then the containment problem has been thoroughly studied (e.g., Wagner and Gorelick, 1989; Lee and Kitadinis, 1991; Shafike *et al.*, 1992; Woldt and Bogardi, 1992; Tiedman and Gorelick, 1993). A review of optimisation and decision analysis for aquifer restoration and contaminant migration-control through pump-and-treat was made in Freeze and Gorelick (1999). These methods also have high potential in the design of regional or local networks for reference level monitoring.

Information transmission methods are based on the entropy as defined by Shannon (1948), on marginal entropy (Shannon and Weaver, 1949), and on the definition of transinformation (Amorocho and Espildora, 1973). Entropy is a measure of uncertainty; marginal entropy is a measure of the uncertainty of one outcome given the knowledge of a second outcome related to the first; transinformation is the reduction of the original uncertainty of one outcome given the knowledge of the second. These methods were applied to water resources with good results (e.g., Amorocho and Espildora, 1973; Caselton and Husain, 1980; Harmancioglu and Yevjevich, 1987; Harmancioglu and Alspaslan, 1992).

The inclusion of temporal information in the design of monitoring networks has been addressed in the geostatistical context as part of the covariance/variogram model by Bastin *et al.* (1984), Lebel *et al.* (1987), Buxton and Pate (1994), and Pardo-Igúzquiza (1998), among others. The resulting models are however, both very complex and problem-specific: the climatological variogram, in the form of a linear function of a time-dependent, space invariant, scaling parameter and a variogram model, in the case of the first two and last references; or a variogram model consisting of a three-dimensional isotropic spherical structure, a one-dimensional zonal linear structure, and a one-dimensional cosine structure, in the case of Buxton and Pate (1994). The sampling strategy may be designed not only to improve the precision of the estimated field, but also to target the areas that exhibit critical

estimated values (Rouhani and Hall, 1988; ASCE, 1990b), higher temporal variability, or special temporal features that distinguish them from the other spatial locations. This limitation in the former models can be overcome by explicitly incorporating a temporal component in the objective function (OF), as is proposed here.

Monitoring network designs is limited by budgetary constraints (except for very special cases), which adds important limitations to the dimension of a particular design with regard to the number of stations to be included. In groundwater monitoring networks these costs are investment and exploration costs. If the design is made from scratch then the number, location, frequency of monitoring and characteristics of the wells can all be included in the design, and the problem is multiobjective: maximizing the information gained, while keeping the investment and exploration costs at a minimum. However, if the objective is to reduce the dimension of an existing monitoring network, then exploration costs can be included in the OF, and the problem is formulated using location as the only variable. Some authors have addressed this problem in the context of variance reduction to optimise rainfall data collection network designs (Bras and Rodríguez-Iturbe, 1976), to assess the effect of altering the size and configuration of a groundwater monitoring network for groundwater levels (Candela *et al.*, 1988), and to establish an optimal network design to estimate the areal averages of rainfall events (Pardo-Igúzquiza, 1998).

When reducing the number of stations of a GMN the problem is combinatorial: which sub-set of stations, π , out of the original set, Π , should be retained? If the number of stations is large then the dimension of the combinatorial problem may be exhaustively intractable. Pardo-Igúzquiza (1998) recently solved this problem for rain gauges using simulated annealing. Rouhani and Fiering (1986) also used variance reduction techniques to determine the number and position of groundwater monitoring stations, and analysed the robustness and resilience of these methods (robustness was defined by Matalas and Fiering (1977) as ‘the insensitivity of a system design to errors, random or otherwise, in the estimates of those parameters affecting design choice’; resilience in the definition of Shannon (1948) is ‘the ability of the system to accommodate surprises and to survive under unanticipated perturbations’). Rouhani and Fiering (1986) found a significant degree of instability in the parameters of the covariance function (parameter space), though this instability had negligible effect on the action space. Similar results were obtained by Shannon (1948). Rouhani and Hall (1988) proposed the incorporation of risk, defined as the weighted sum of the expected value and the estimation variance, in order to correct the ‘blindness’ of the estimation variance to extreme values. Longer reviews on this theme may be found in Loaiciga *et al.* (1992), Dixon and Chiswell (1996), and Harmancioglu *et al.* (1999).

In this article the problem of reducing a GMN dimension is treated in the context of variance reduction, coupled with a measure of temporal redundancy, such that important local temporal variabilities are detected, and exploration costs in the

form of the time taken to travel between stations. The term *data quality* is used when referring to both spatial accuracy and temporal redundancy, and *sampling time* when referring to the sum of monitoring and travel times. An objective function was built to incorporate these four factors.

The selection of the new GMN design was made by a cost-benefit analysis of the gain in data quality and sampling times. The OF was optimised for several π values with simulated annealing (SA), and the optimal π value obtained.

Two new developments in GMN optimisation are introduced in this article: (i) a new definition of temporal redundancy; (ii) the minimisation of sampling times by incorporating them directly in the OF.

The article is organized in the following way: in the second section an objective function model is proposed; in the third section the simulated annealing algorithm proposed to solve the optimisation problem is described, and its application is discussed; the fourth section presents a hypothetical case-study to test the method; the results of the application are presented and discussed in the fifth section; finally, in the sixth section the most relevant conclusions are drawn.

2. Objective Function

The new GMN size and spatial distribution must assure that both spatial and temporal variabilities are correctly included in the new design, and that exploration costs (sampling time) are minimized.

It is better to consider all the information collected, rather than using statistical transformations of the data. Common statistical variables are: maximums, minimums, percentiles, average values, variance, or more evolved transformations like projection in factorial spaces and similarity (or correlation) coefficients. Most of these statistical transformations have the advantage of reducing a large amount of data to a treatable number of parameters. Despite their particular advantages, these methods tend to enhance some particular features of the data, while disregarding others. One way to circumvent this *blindness* is to use raw data whenever possible. A mathematical time series function that makes no statistical transformation of the raw data is proposed and used in this article. The function is a measure of data redundancy and is used to obtain better spatial distributions of monitoring stations.

Spatial accuracy, temporal redundancy, monitoring times for each station and total travel time are incorporated in a single OF and subjected to minimisation. Simulated annealing (SA) is proposed to optimise the OF.

2.1. SPATIAL ACCURACY

Monitoring network optimisation for geo-variables, such as in areal mean rainfall events, was the subject of early works by Rodríguez-Iturbe and Mejía (1974a, b), in which they considered the spatial and temporal variability of mean rainfall. The

variance of mean rainfall was calculated as the product of point process variance, a reduction factor given by sampling in time (dependent only on the correlation in time and length of the time series), and a reduction factor given by sampling in space (dependent on the spatial correlation structure, the sampling geometry and the number of stations). These authors studied random and stratified random sampling schemes, and obtained an abacus for different correlation functions, the number of stations, and area of the region. Lenton and Rodríguez-Iturbe (1974) also considered the density and location of the stations. Bras and Rodríguez-Iturbe (1975) continued the latter's work, comparing former results for different point variances, the covariance function and covariance function parameter. In the same formalism, Bras and Rodríguez-Iturbe (1976) included the cost associated with each station to help choose the best set (number and position) of stations with least cost and least mean rainfall variance. The methods based on the reduction of the estimation errors variance are known as variance-reduction techniques.

At this stage the variance-reduction method is developed using geostatistical nomenclature. Within the geostatistical formalism, field data are held to be the result of random processes of regionalized variables, i.e., of random variables with space coordinates, with some spatial covariance. Regionalized variables are continuous in space, and therefore not completely random, but at the same time it is not possible to model them by means of a deterministic function (or spatial process). They therefore result from deterministic and stochastic processes (Matheron, 1970), incorporating the notion of uncertainty in the conception of inference models or in the simulation of variables (Matheron, 1970; David, 1977; Journel and Huijbregts, 1978). ASCE (1990a, b) gives a thorough review on the use of geostatistics for mapping and sampling design.

The values of $z(x)$ at the sampled points in the field can be considered as realizations of a set of random variables $Z(x)$ in a field Γ . A set of random variables $Z(x_i)$ defined in a field Γ is a random function $Z(x)$:

$$Z(x) = \{Z(x_i)\}_{x_i \in \Gamma}$$

$z(x_i)$ is therefore a realization of a random variable, and the latter a realization of a random function.

As field data are usually scarce, it is not possible to calculate the distribution function, but for practical applications only the first two moments are needed. Some restrictions with respect to stationarity are needed. The most common theory considers that the distribution function is invariant by translation. As, strictly speaking, the restrictive hypotheses are applied only to the first two moments, it is required that these exist and that they are independent of space coordinates (second order stationarity), otherwise the spatial covariance of the variable Z depends solely on the separation vector \mathbf{h} (in modulus and direction). In this case only the spatial

increments have to be stationary (intrinsic stationarity). If these increments are made at step h , then the resulting expression is called the variogram:

$$\gamma(h) = \frac{1}{2} E \{ [Z(x+h) - Z(x)]^2 \} . \quad (1)$$

The matrix system used to compute the kriging weights has to be invertible. For this property to be true, all the covariances of the system must be defined from positive definite functions (see e.g., Matheron, 1970, p. 54). The variogram has tended to replace covariance, mainly because of the less restrictive intrinsic hypothesis. In most practical cases this less demanding restriction on the random function model will not change the results (Deutsch and Journel, 1992).

To estimate the average value of a variable in an area A from values at locations x_α , $Z(x_\alpha)$, inside or outside the area, this value is equal to

$$V = \frac{1}{A} \int_A Z(x) dx . \quad (2)$$

A linear estimation of V can be obtained from n data points by

$$V = \sum_{i=1}^n \kappa_i \cdot Z(x_i) , \quad (3)$$

which is unbiased if the sum of the weights κ is one. This is a common requirement in several methods and also in kriging. This method is chosen because values of κ are determined so as to minimize the variance of the error of estimation. The kriging system is defined by (Journel and Huijbregts, 1978):

$$\begin{cases} \sum_{k=1}^n \kappa_k \cdot \gamma(h_{ij}) + \mu = \gamma(h_{iA}) & i = 1, \dots, n \\ \sum_{k=1}^n \kappa_k = 1 , \end{cases} \quad (4)$$

where μ is the Lagrange parameter, and $\gamma(h_{iA})$ is the average variogram between the point i and the area A when one extreme of the vector h is fixed in x_i and the other extreme describes the area A independently. The average variogram is then:

$$\gamma(h_{iA}) = \frac{1}{A} \int_A \gamma(x_i, u) du . \quad (5)$$

Approximated numerically by

$$\gamma(h_{iA}) \approx \frac{1}{M} \sum_{k=1}^M \gamma(x_i, x_j) \quad x_j \in A , \quad (6)$$

with M number of points used to discretise A . The estimation variance is expressed by (Journel and Huijbregts, 1978):

$$\sigma_E^2 = \sum_{i=1}^N \kappa_i \cdot \gamma(h_{iA}) - \gamma(h_{AA}), \quad (7)$$

and

$$\gamma(h_{AA}) \approx \frac{1}{M^2} \sum_{i=1}^M \sum_{k=1}^M \gamma(x_i, x_k) \quad x_i, x_k \in A. \quad (8)$$

The estimation variance is a measure of the estimation accuracy of V . Because σ_E^2 only depends on the data points' location, and once a variogram model is defined, it is possible to change data location and calculate the estimation variance again. The spatial arrangement of points that minimizes σ_E^2 has the lowest estimation error, and therefore best reflects the spatial correlation introduced into the variogram model.

2.2. TEMPORAL REDUNDANCY

All time series are considered synchronous and complete, either because data were collected at the same time, or because the necessary interpolations were made to synchronize data and fill in the gaps. Time events can therefore be handled as realizations of random functions, $Y_i(m)$, $i = 1, \dots, L_{FIXED} + L_{EXP}$; $m = 1, \dots, D$, (with L_{FIXED} being the number of fixed stations, i.e., those that are to be included in all solutions; L_{EXP} the dimension of the subset of stations to be included in the new design; D length of the time series vectors). $Y_i(m) \in X \in \chi$, with X being the current solution, and χ the solution space.

If the time series are not synchronous and/or complete then it is proposed that: (i) data should be interpolated, or (ii) a measure of temporal correlation be used, in either the frequency or correlation domains.

The sum of differences between time series, S , is used to evaluate redundancy: large values of S indicate that the series are very different, while smaller values indicate the opposite. S is an approximation to the π -integral of time series functions.

The series are first normalized by subtracting the average value, $\bar{Y}_i : Y_i^0(m) = Y_i(m) - \bar{Y}_i$. Then the summation of the difference between the series is made for all times, with the possibility of shifting one time series in relation to the others by the time value Δ so that the sum is the lowest, i.e., when summing the difference between series $Y_i^0(m)$ with series $Y_k^0(m)$ $k \neq i$, only the minimum values are used. This time translation should guarantee that time series are compared in-phase. Minimum temporal redundancy implies maximum S :

$$S(n) = \text{Max} \left\{ \sum_{\substack{i=1 \\ k=i+1}}^{L\text{FIXED}+L\text{EXP}-1} \right. \\ \left. \text{Min} \left\{ \frac{\sum_{m=1}^{D-d} Y_i^0(m) - Y_k^0(m+d)}{D-d}, 0 \leq d \leq +\Delta \right\} \right\} . \quad (9)$$

2.3. EXPLORATION COSTS

The costs associated with maintaining a GMN are considered to be dependent only on the time needed to measure at each station, in the form of equal or different times for measuring per station, and the time needed to travel through the whole network along the shortest path. Other costs not included in this formulation are, e.g., maintenance costs and the cost associated with the non-detection of a pollution event (as considered, e.g., in Massmann and Freeze, 1987a, b; Wagner and Gorelick, 1987; Rouhani and Hall, 1988; Meyer *et al.*, 1994). The sum, UC, of individual measuring times, uc_i , $i = 1, \dots, \pi$, is used to calculate the OF model,

$$UC = \sum_{i=1}^{\pi} uc_i . \quad (10)$$

The determination of minimum total travel time (TC) is a well-known combinatorial problem known as the travelling salesman problem (TSP), solved here with the algorithm proposed by Carpaneto *et al.*, (1995a, b). The optimal TC obtained by the algorithm is used to calculate the OF.

2.4. OBJECTIVE FUNCTION MODEL

The network of π stations with higher spatial accuracy, lower temporal redundancy, and lower exploration costs (times), implies the minimization of the following objective function,

$$\text{OF} = w_1 \cdot \frac{\sigma_E^2}{\text{Max}(\sigma_E^2)} + w_2 \cdot \left[1 - \frac{S}{\text{Max}(S)} \right] + w_3 \cdot \frac{UC}{\text{Max}(UC)} + w_4 \cdot \frac{TC}{\text{Max}(TC)} \quad (11)$$

Conditional to:

$UC + TC \leq 7 \text{ hours}$

available data

$\text{Max}(\sigma_E^2)$, $\text{Max}(S)$, $\text{Max}(UC)$, and $\text{Max}(TC)$ are the maximum values found during the optimisation process, and are needed to make the OF dimensionless and problem-independent. It is assumed that the monitoring takes only one working-day of seven hours. This can be easily altered if necessary, however. The weights

$w_j, j = 1, \dots, 4$, are such that $\sum w_j = 1$. This may be useful if one wishes to give more emphasis to the quality of the data collected, by giving higher weights to the first two variables, or to control the exploration costs more stringently, by giving higher weights to the two remaining variables. In the example given here, equal weights ($= 1$) are used for all variables.

3. Solving the Model

Reducing a monitoring network of dimension Π to a network of smaller dimension, π , is a problem for which the number of possible combinations is given by $\Pi!/(\pi! \cdot (\Pi - \pi)!)$. However, if the dimension of the original network is large, there may likewise be a very large number of combinations, and computing the OF exhaustively would require much more time than is usually available. Therefore a more efficient method has to be used in order to obtain a good quality solution (in the sense that it may not be the global optimum) in a reasonable amount of time.

It is proposed to solve the problem by replacing one station at a time, evaluating the result, keeping the station if it reduces the OF or if the result fulfils a probabilistic criterion (the Metropolis criterion), and rejecting the station otherwise. The iterative process of replacing the stations and analysing the fulfilment of the Metropolis criterion is a crucial part of the simulated annealing algorithm. The algorithm looks at each iteration for the cost of a given set of stations (a feasible solution), and it may be that a station that is rejected in one iteration, with a particular combination of stations, is accepted in a later stage with a different combination (different solution). Therefore an efficient search of the solution space has to be ensured. A global optimal solution would be found if the entire space was searched. However, in practical applications, if the number of combinations is very large, a global search would take far too much time. Algorithms that do not guarantee a global search can only provide optimal local solutions. This is the case with the simulated annealing (SA) implementation used here.

SA is one of the threshold algorithms included in the class of local search algorithms. The other two, as defined by Aarts and Korst (1990), are: iterative improvement, where only OF-reducing neighbours are accepted; and threshold accepting, where a certain deterministic non-increasing threshold sequence is used, allowing neighbour solutions with larger OF to be accepted, though to a limited extent. This is because the threshold value is fixed and always decreasing, with a very rigid control on the size of the cost difference. Simulated annealing uses a more flexible control on the threshold values, allowing transitions out of a local minimum at nonzero temperatures.

SA was first introduced by Kirkpatrick *et al.* (1983) as an algorithm to solve very well-known combinatorial optimisation problems, reducing the risk of falling prematurely into local minima (or metastable solutions) common to iterative improvement methods, because they tend to accept only solutions that lower the

OF, with fast quenching from high temperatures to temperatures near zero. These authors proposed the use of the Metropolis (Metropolis *et al.*, 1953) procedure from statistical mechanics. This procedure generalizes iterative improvement by incorporating controlled uphill steps (to worse solutions). The procedure states the following: consider one small random change in the system at a certain temperature; the change in the objective function is ΔOF ; if $\Delta OF \geq 0$, then the change in the system is accepted and the new configuration is used as the starting point for the next step; if $\Delta OF < 0$ then the probability that the change is accepted is determined by $P(\Delta OF) = \exp(-\Delta OF/k_b T)$; a random number uniformly distributed in the interval (0,1) is taken and compared with former probability; if this number is lower than $P(\Delta OF)$ then the change is accepted. The control parameters $k_b T$ are replaced by the parameter t (also called temperature), to avoid using the Boltzman constant, k_b , which would have no meaning in this context.

The SA algorithm runs in the following way: (i) the system is *melted* at a high temperature (initial temperature, t_1); (ii) the temperature is decreased gradually until the system *freezes* (no further OF change occurs); (iii) at each iteration the Metropolis procedure is applied.

The generic SA algorithm for minimisation, considering a neighbourhood structure N , a solution space χ , and an objective function OF has the following pseudo-code.

```

Select an initial solution  $X_{best}$ ;
Select an initial temperature  $t_1 > 0$ ;
Select a temperature reduction factor;
Repeat
    Repeat      Randomly select  $X \in N(X_{best})$ ;
                 $\delta = OF(X) - OF(X_{best})$ ;
                if  $\delta < 0$  then
                     $X_{best} = X$ 
                else
                    generate random  $z$  uniformly in (0,1);
                    if  $z < \exp(-\delta/t)$  then  $X_{best} = X$ ;
    Until  $iterations = max\_iterations$ 
    Set  $t = \alpha(t)$ ;
Until stopping condition = true;
 $X_{best}$  is the optimal solution found.

```

Once a feasible solution is generated it is passed to the OF calculation module (Figure 1). Here, two routines are internal (Redundancy and Monitoring time) and the other two are external, and were developed elsewhere. The following data is read: (i) (x, y) coordinates and data values matrix into the kriging routine; (ii) time series values matrix into the redundancy routine; (iii) monitoring times per station

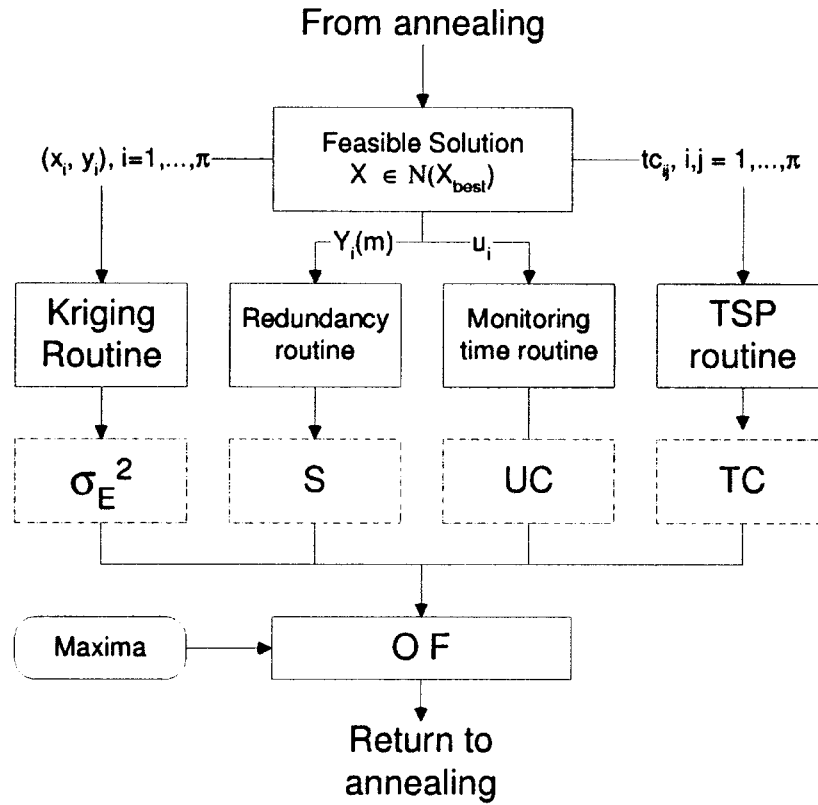


Figure 1. Objective function computing module.

matrix into the monitoring time routine; (iv) travel distance matrix into the TSP routine.

In order to speed-up the process several improvements have been proposed, particularly, limiting the number of iterations at each temperature, i.e., defining the *max_iterations* number. The dimension of the Markov chain has been proposed as a function of the dimension of the problem (Kirkpatrick *et al.*, 1983): temperature is maintained until 100Π solutions (iterations), or 10Π successful solutions have been tested, whichever comes first. Π stands for the number of variables (stations) in a problem. These authors also proposed that the annealing is stopped (stopping criterion) if after three consecutive temperatures the number of acceptances is not achieved. Alternatively, if the average value of the OF does not change after a pre-established number of temperature decreases (R_{STOP}), then the annealing is stopped. Along with these dynamic criteria, a static one may be used to halt the process when a minimum temperature, t_{min} , is reached. This last criterion will guarantee that the annealing stops if none of the dynamic criteria is fulfilled, even before the total number of iterations is attained. In the present algorithm, both the dynamic and the static criteria were implemented.

Cunha and Sousa (1999) proposed the following expression to calculate the initial temperature, t_1 ,

$$t_1 = -\frac{b \cdot \text{OF}_0}{\ln a}, \quad (12)$$

where OF_0 is the cost of the initial configuration, a is the elasticity of acceptance, and b is the amount of dispersion around the cost of the initial solution. Prior runs of the model determine the last parameter. The elasticity of acceptance represents the probability of accepting a solution worse than the initial one. The initial temperature determined by (Equation (12)) is such that there is a probability a of accepting solutions that are $b\%$ worse than the initial solution.

Temperature is usually decreased at a constant rate (cooling factor), α , such that after s temperature decreases the temperature is $t_s = t_1 \cdot \alpha^s$. The two stopping criteria: t_{\min} and number of temperature decreases are complementary because it is easy to calculate the minimum temperature attained if t_1 and α are known.

A specific computer code in FORTRAN was developed by the authors to solve GMN reduction problems (MINCOST). The code incorporates two routines developed by others, namely the kriging routine (Deutsch and Journel, 1992), and the solution for the travelling salesman problem (Carpaneto *et al.*, 1995a, b). The code was thoroughly tested and validated.

4. Case Study

In order to test the method, a grid of 32 monitoring stations randomly distributed in a square grid of eight by eight spatial units was created (Figure 2). Monitoring stations numbers 1 and 2 represent wells for water supply and are therefore included in all the solutions. Π is therefore equal to 30. Time series of generic geo-variables for all stations were calculated using common mathematical functions, purely deterministic and with normally distributed errors. The functions were selected based on empirical judgment and experience, attempting to simulate the behaviour of water quality variables. Some of the resulting variables are non-homocedastic, thus reflecting a feature common to variables such as redox potential and electrical conductivity when the scale of measure has to be changed between consecutive observations. The unit of the generic geo-variable is identified by un. The time series equations are presented in Table I.

The intention was then to choose the π monitoring stations that should be added to stations 1 and 2 in order to obtain a network that has optimal exploration costs, best preserves the temporal features of the data, and has optimal spatial accuracy.

The first goal was to determine the number of stations to be included in the new design; the second goal was to identify the stations in the new optimal design.

TABLE I
Time series equations

Series number	Equation: $x =$
1	$\cos(\text{time}) + 0.54 \Delta \text{time}$
2	0.12 time
3	$\sin(\text{time})$
4	$\cos(\Delta \text{time} + \text{time})$
5	$\cos(2\Delta \text{time} + \text{time})$
6	$\log(\text{time}) - \log(\Delta \text{time})$
7	$\exp[-\text{time}/(\text{end time} - \text{start time})]$
8	1
9	$\cos(0.3 \text{ time} + 3\Delta \text{time})$
10	$\text{atan}(\text{time})$
11	$\text{atan}(\text{time} + 1.5\Delta \text{time})$
12–22	Repeat Equations (1)–(11) + 20% normally distributed error around the mean
23–32	Repeat Equations (1)–(10) + 40% normally distributed error around the mean

4.1. RESULTS

4.1.1. Cost-benefit analysis

Eight different subset cardinalities of $\pi + 2$ stations were tested, {13, 15, 17, 19, 21, 23, 25, 27}, and a cost-benefit analysis was performed for the gain in accuracy and temporal representativeness and the increase in exploration costs (time needed to sample).

As expected, the estimation variance decreased as new stations were added (Table II) (indicating an increase in spatial accuracy), but the gain in spatial accuracy was much more significant up to the 23rd station, with each new station producing an average increase in spatial accuracy of 8.1%; after the 23rd station the spatial accuracy only increased 0.5% each time a new station was added.

It should be noticed here that the interpretation of data redundancy (S) is different when considered in the context of cost-benefit analysis or in the context of network optimization for a pre-defined π number. In the first case $S(\pi + 2)$ represents the total amount of information in the time series, and $S(\pi_2) - S(\pi_1)$, with $\pi_1 < \pi_2$, is the marginal S, or the amount of information added by the $\pi_2 - \pi_1$ stations. In the second case no marginal S is calculated, and S is always the total S, calculated for different $\pi + 2$ values.

The gain in temporal representativeness (marginal S) increased every time a new station was added, but the average rate of increase was almost four times higher up to the 19th station than it was for the remaining stations (9.8% against

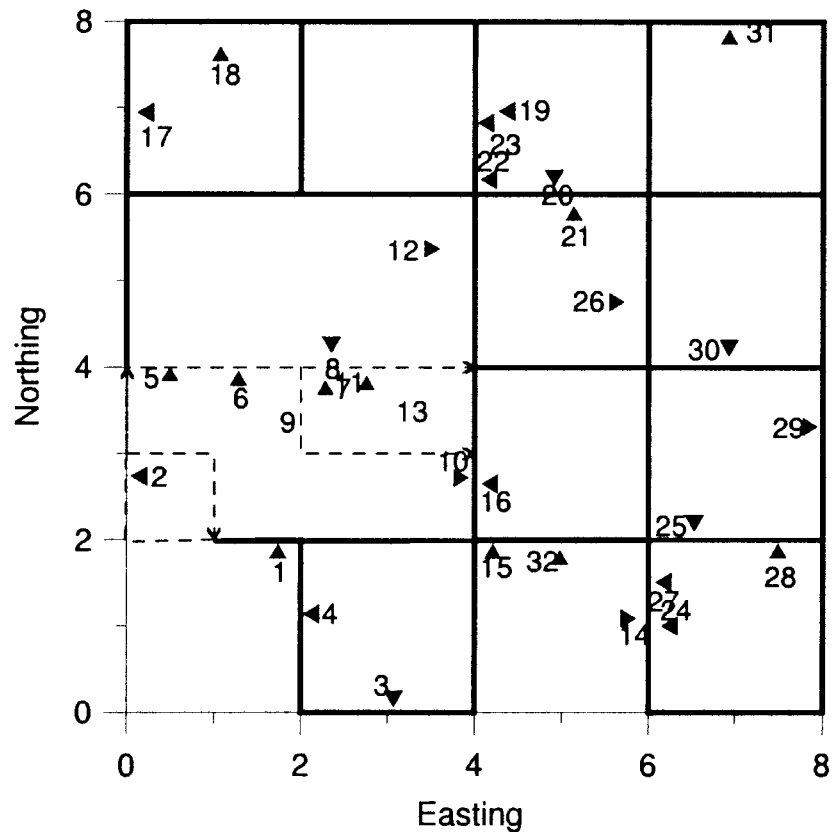


Figure 2. Location of the monitoring stations and routes connecting them: bold line for two-way.

2.5%). Figure 3a shows the accuracy and representativeness curves with the fitted trend lines drawn as interrupted straight lines.

Both monitoring and travel times (UC and TC) increased polynomially (2nd degree) with the number of stations, with UC having a faster growth rate than TC. One interesting feature is the stabilization of travel times after the 23rd station; a similar stabilization is found for the monitoring time (UC) after the 21st station, but in this case UC increases again after the 23rd. As a result the increase in total sampling time shows a polynomial behaviour up to the 21st station, after which it becomes linear. Figure 3b shows the variation of UC, TC and UC + TC, with the number of stations and the adjusted second-degree polynomials.

These results show an interesting data quality/sampling time structure: up to the 23rd station, though data quality increases linearly, sampling times increase polynomially; after the 23rd station the increase in data quality is much lower and the sampling times still increase, but at a lower rate. This indicates that a monitoring network with 23 stations would be the optimal solution.

TABLE II
Optimization results for the eight π values. The original GMN results are also show

Number of stations	OF	σ_E^2 (un ²)	S (un)	UC (h)	TC (h)	UC + TC
13	1.885	0.09788	204.9	1.841	1.327	3.168
15	1.966	0.09808	264.3	2.190	1.362	3.552
17	1.915	0.08715	297.6	2.563	1.406	3.969
19	2.032	0.08827	349.0	2.935	1.482	4.418
21	2.053	0.07273	417.0	3.635	1.692	5.326
23	2.044	0.06209	424.3	3.634	2.036	5.670
25	2.021	0.06336	464.8	4.147	1.982	6.129
27	2.119	0.06081	481.3	4.485	1.999	6.483
32	–	0.05937	550.4	5.990	2.704	8.694

Sampling times cost structure may be investigated by analysing the relative contribution of UC and TC to the total time. Figure 4 shows the relative contribution of UC to the total sampling time. UC is always the most important cost factor, and its importance increases with the dimension of the network, at least up to the 21st station, after which the relative weight tends to stabilise around 70%. The jump at the 23rd station is justified by the need to introduce new stations in locations with difficult access, for which the new path is longer; the subsequent increase indicates that the stations added after the 23rd can be connected by the new-found path, and UC becomes much more important again. Therefore, new gains in efficiency (reduction of total sampling time) should be sought by reducing the measuring times, μ_i .

The selection of stations show high congruency for the different $\pi + 2$ values (see Figure 5): (i) four of the stations are not included in any design (7, 9, 10 and 13); (ii) four of the stations are only included for very high $\pi + 2$ values (≥ 25) (5, 6, 11, 25). The interpretation is complex due to the presence of four variability factors, but it is still possible to observe the following: the stations mentioned make a small contribution to the increase in spatial accuracy, but, on the other hand, they tend to increase the total travel time (their location being on one-way roads). Furthermore, the stations involve low marginal Ss (see Figures 3a and b).

4.2. SELECTED MONITORING NETWORK

The selected $\pi + 2$ station GMN is now analyzed, but equivalent analyses could be made for different π values, not included due to space limitations. Figure 6 shows the OF, data quality and sampling times convergence curves. The OF starts in the first iterations at high temperatures, with values close to 2.7, indicating that it is still

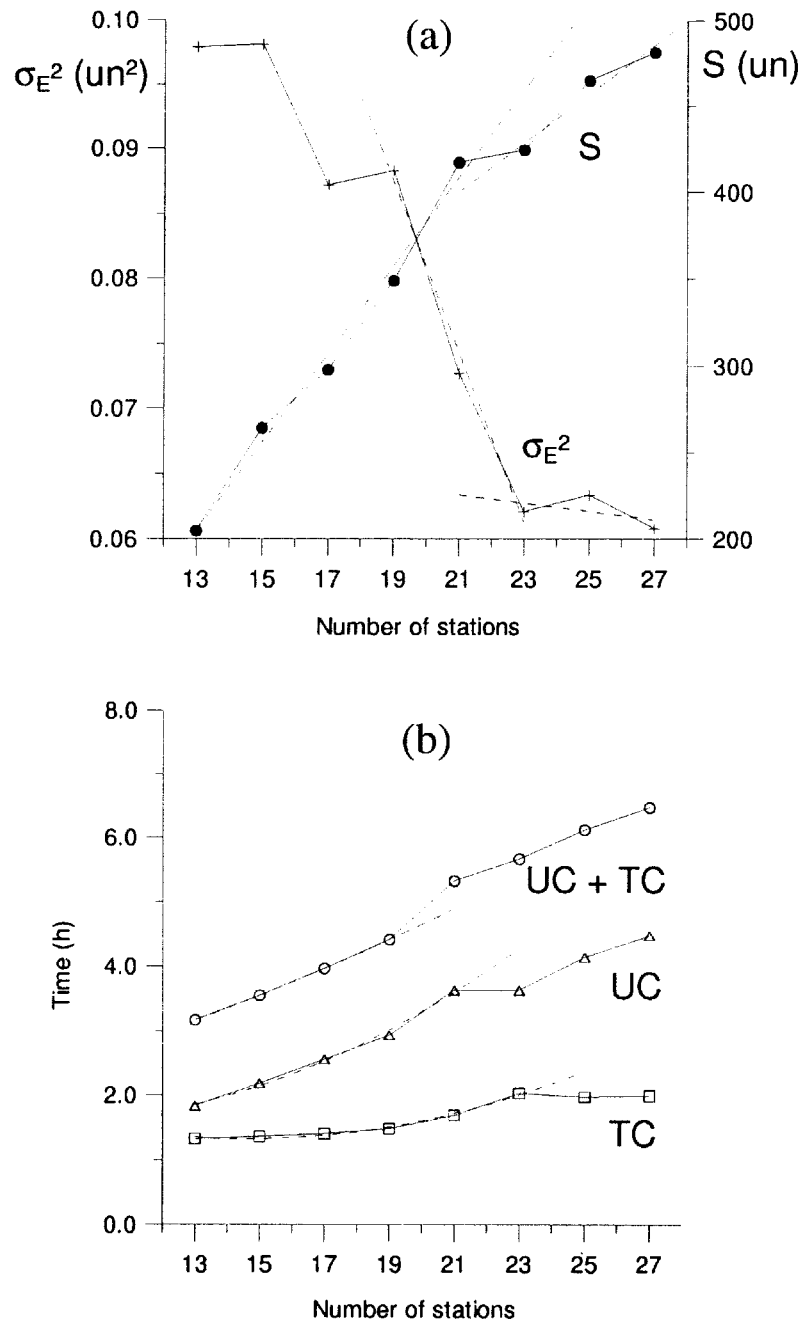


Figure 3. Cost-benefit analysis for increasing number of stations: a) gain in spatial accuracy and temporal representativeness; b) Time necessary to sample.

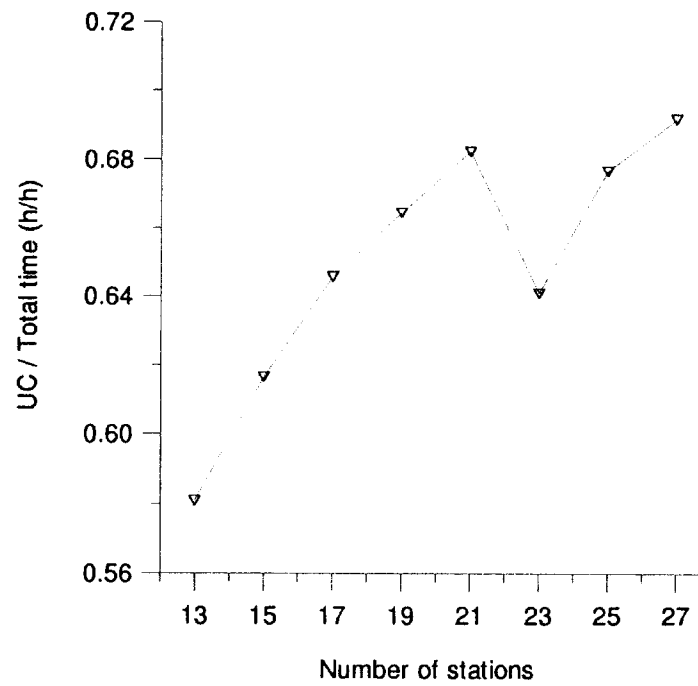


Figure 4. Relative contribution of UC to the total sampling time.

Station's number: 3-32

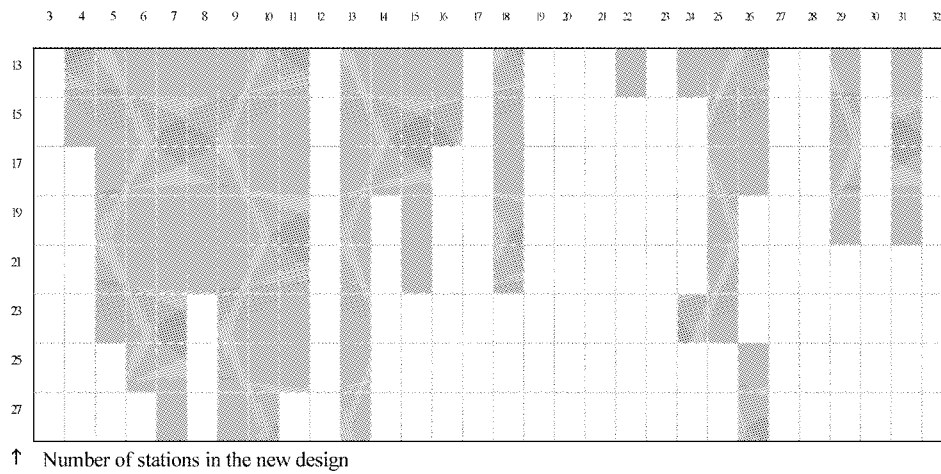


Figure 5. Stations excluded in the new designs (in gray).

far from its minimum. This is clearly visible in Figure 6, as the four OF variables slowly converge towards their optimal values. At sufficiently low temperatures a *frozen* system is achieved and the optimal (local) solution is reached. The solution GMN is shown in Figure 7.

The resulting $\pi + 2 = 23$ stations GMN is 28 % smaller than the original one, and results in a reduction of 3.0 hr in sampling time (35%). At the same time, only 20% of spatial accuracy (0.0119 un^2) and 22.9% of total time series information (126.1 un) was sacrificed.

5. Conclusions

GMN optimisation problems are particularly difficult to solve because of discontinuous integer-related nonconvexity (as a monitoring station is either included in the new design or not, the feasible region is discontinuous) and continuous nonconvexity of the interpolation error variance surface (due to the presence of many local minima). Some solvers can handle one or the other type of nonconvexity, but the combination of both requires heuristic methods. Because the aim of this article was to test the objective function model, no comparison of algorithms was made, rather, simulated annealing was selected for its good results in other optimisation problems. Furthermore, it has already been tested for monitoring networks design against other heuristic algorithms, and it has outperformed the them (TABU search, genetic algorithms, and sequential exchange search algorithms) (Lee and Ellis, 1996).

In many instances it is neither possible nor advisable to incorporate statistics-based frequency methods in the optimisation models due to the short length of time-series, missing data, and the bias introduced in the data by the empirical selection of locations and frequencies of the existing monitoring networks. An automatic frequency/location optimisation will, in these cases, seriously compromise the quality of the solution. A better alternative is to establish the frequency beforehand, after judicious analysis of the available time series and the hydraulic properties of the aquifer (e.g., in karst aquifers frequency must, in most cases, be higher than in porous aquifers). The location and number of stations are then selected afterwards. This is the case of the GMN optimisation problem solution described here for reference level/compliance monitoring with a pre-set sampling frequency.

The method proposed in this article seems to have a good applicability for reference level ground water monitoring networks when the main spatial variability and behaviour of time series should be preserved, but exploration costs are to be optimised as well. Due to the definition of temporal redundancy, which is calculated in the data space rather than covariance space (Amorocho and Espildora, 1973; Caselton and Husain, 1980; Harmancioglu and Yevjevich, 1987; Harmancioglu and Alpaslan, 1992), there is a tendency to overburden time series with higher values

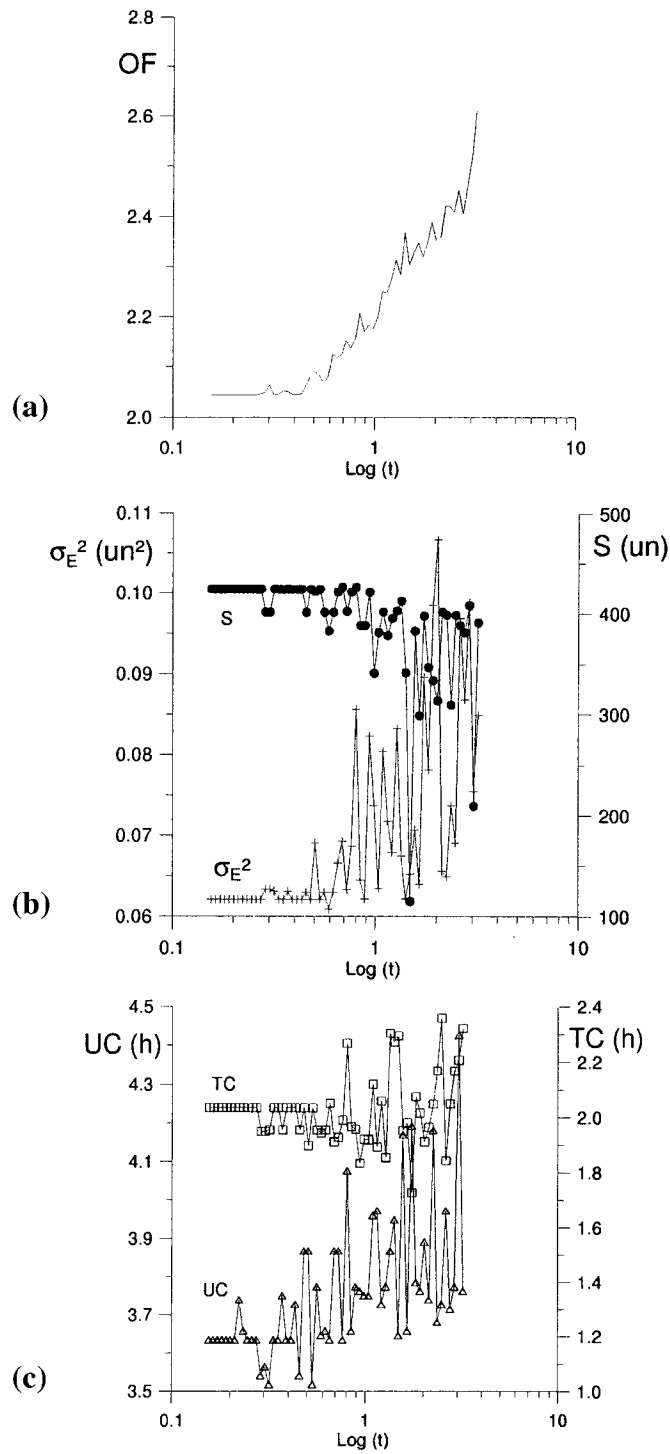


Figure 6. Optimization results for a monitoring network of 23 stations. (a): Objective function; (b) data quality variables; (c) sampling times.

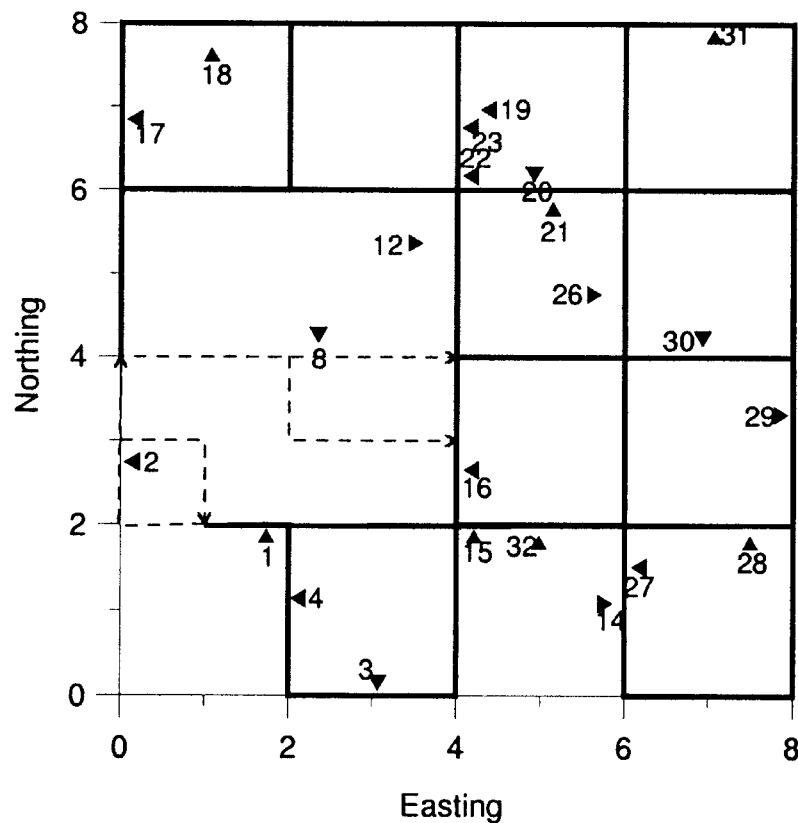


Figure 7. Resulting monitoring network with 23 stations.

and/or higher variability. This feature is, however, welcome in most situations, and is an accepted penalty for not having to consider particular variable or multivariate statistical distributions.

The results showed that it is possible to eliminate some stations from a GMN and have a relative lowering of costs greater than the relative reduction of GMN dimension and collected data quality.

References

- Aarts, E. and Korst, J.: 1990, *Simulated Annealing and Boltzman Machines*, John Wiley & Sons, New York.
- Amorocho, J. and Espildora, B.: 1973, 'Entropy in the assessment of uncertainty of hydrologic systems and models', *Water Resour. Res.* **9**(6), 1511–1522.
- ASCE: 1990a, 'Review of geostatistics in geohydrology. I. Basic concepts', *J. Hydraulic Eng.* **116**(5), 612–632.
- ASCE: 1990b, 'Review of geostatistics in geohydrology. II. Applications', *J. Hydraulic Eng.* **116**(5), 633–658.

- Bastin, G., Lorent, B., Duque, C. and Gevers, M.: 1984, 'Optimal estimation of the average rainfall and optimal selection of rain-gauge locations', *Water Resour. Res.* **20**(4), 463–470.
- Bras, R. L. and Rodríguez-Iturbe, I.: 1976, 'Network design for the estimation of areal mean of rainfall events', *Water Resour. Res.* **12**(6), 1185–1195.
- Bras, R. L. and Rodríguez-Iturbe, I.: 1975, *Rainfall-runoff as Spatial Stochastic Processes: Data Collection and Synthesis*, Ralph M. Parsons Lab. for Water Resources and Hydrodynamics, Dep. of Civil Eng., Mass. Inst. of Technology, Cambridge, U.S.A.
- Buxton, B. E. and Pate, A. D.: 1994, 'Joint Temporal-Spatial Modeling of Concentrations of Hazardous Pollutants in Urban Air', in R. Dimitrakopoulos (ed.), *Geostatistics for the Next Century*, Kluwer Academic Publishers, The Netherlands, pp. 75–87.
- Candela, L., Olea, R. A. and Custodio, E.: 1988, 'Lognormal kriging for the assessment of reliability in groundwater quality control observation networks', *J. Hydrol.* **103**, 67–84.
- Carpaneto, G., Dell'Amico, M. and Toth, P.: 1995a, 'Algorithm 750: CDT: A subroutine for the exact solution of large-scale, asymmetric traveling salesman problems', *ACM Trans. Math. Soft.* **21**(4), 410–415.
- Carpaneto, G., Dell'Amico, M. and Toth, P.: 1995b, 'Exact solution of large-scale, asymmetric traveling salesman problems', *ACM Trans. Math. Soft.* **21**(4), 394–409.
- Caseltan, W. F. and Husain, T.: 1980, 'Hydrologic networks: Information transmission', *J. Water Resour. Plann. Manag. Div.* **106**(WR2), 503–519.
- Chilès, J.-P. and Delfiner, P.: 1999, *Geostatistics, Modeling Spatial Uncertainty*, John Wiley & Sons, U.S.A.
- Cunha, M. C. and Sousa, J., 1999, 'Water distribution network design optimization: Simulated annealing approach', *J. Water Resour. Plann. Manag.* **July/August**, 215–221.
- David, M.: 1977, *Geostatistical Ore Reserve Estimation*, Elsevier Publ., Amsterdam, The Netherlands.
- Delhomme, J. P.: 1978, 'Kriging in the hydrosociences', *Adv. Water Resour.* **1**(5), 251–266.
- Deutsch, C. V. and Journel, A. G.: 1992, *GSLIB, Geostatistical Software Library and User's Guide*, Oxford University Press, New York, U.S.A.
- Dixon, W. and Chiswell, B.: 1996, 'Review of aquatic monitoring program design', *Water Resour.* **30**(9), 1935–1948.
- Freeze, R. A. and Gorelick, S. M.: 1999, 'Convergence of stochastic optimization and decision analysis in the engineering design of aquifer remediation', *Ground Water* **27**(6), 934–954.
- Harmancioglu, N. and Alspaslan, N.: 1992, 'Water quality monitoring network design: A problem of multi-objective decision making', *Water Resour. Bull.* **28**(1), 179–192.
- Harmancioglu, N. and Yevjevich, V.: 1987, 'Transfer of hydrologic information among river points', *J. Hydrol.* **91**, 103–118.
- Harmancioglu, N. B., Fistikoglu, O., Ozkul, S. D., Singh, V. P. and Alspaslan, M. N.: 1999, *Water Quality Monitoring Network Design*, Kluwer Academic Publishers, The Netherlands.
- Journel, A.: 1987, *Geostatistics for the Environmental Sciences, An Introduction*, Applied Earth Sciences Department, Stanford University, California, U.S.A.
- Journel, A. and Huijbregts, Ch.: 1978, *Mining Geostatistics*, Academic Press, New York, U.S.A.
- Kirkpatrick, S., Gellat Jr., C. D. and Vecchi, M. P.: 1983, 'Optimization by simulated annealing', *Science* **220**, 671–680.
- Lebel, T., Bastin, G., Obled, C. and Creutin, J. D.: 1987, 'On the accuracy of areal rainfall estimation: A case study', *Water Resour. Res.* **23**(11), 2123–2134.
- Lee, S.-I. and Kitadinis, P. K.: 1991, 'Optimal estimation and scheduling in aquifer remediation with incomplete information', *Water Resour. Res.* **27**(9), 2203–2217.
- Lee, Y.-M. and Ellis, J. H.: 1996, 'Comparison of algorithms for nonlinear integer optimization: Application to monitoring network design', *J. Environ. Eng.* **June**, 524–531.

- Lenton, R. L. and Rodríguez-Iturbe, I.: 1974, *On the Collection, the Analysis, and the Synthesis of Spatial Rainfall Data*, Ralph M. Parsons Lab. for Water Resources and Hydrodynamics, Dep. of Civil Eng., Mass. Inst. of Technology, Cambridge, U.S.A.
- Loaiciga, H. A., Charbeneau, R. J., Everett, L. G., Fogg, G. E., Hobbs, B. F. and Rouhani, S.: 1992, 'Review of ground-water quality monitoring network design', *J. Hydraulic Eng.* **118**(1), 11–37.
- Massmann, J. and Freeze, R. A.: 1987a, 'Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1. Methodology', *Water Resour. Res.* **23**(2), 351–367.
- Massmann, J. and Freeze, R. A.: 1987b, 'Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 2. Results', *Water Resour. Res.* **23**(2), 368–380.
- Matalas, N. C. and Fiering, M. B.: 1977, 'Water Resource System Planning', in *National Research Council, Climate, Climatic Change and Water Supply*, National Academy of Sciences, Washington, D.C., pp. 64–72.
- Matheron, G.: 1970, 'La Théorie des Variables Régionalisées, et Ses Applications', *Les Cahiers Du Centre De Morphologie Mathématique De Fontainebleau*, **Fascicule 5**.
- McBratney, R. W., Webster, R. and Burgess, T. M.: 1981a, 'The design of optimal sampling schemes for local estimation and mapping of regionalized variables – I', *Comp. Geosc.* **7**, 331–334.
- McBratney, R. W., Webster, R. and Burgess, T. M.: 1981b, 'The design of optimal sampling schemes for local estimation and mapping of regionalized variables – II', *Comp. Geosc.* **7**, 335–336.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: 1953, 'Equation of state calculations by fast computing machines', *J. Chem. Phys.* **21**, 1087–1092.
- Meyer, P. D. and Downey, B.: 1988, 'A method for locating wells in a groundwater monitoring network under conditions of uncertainty', *Water Resour. Res.* **24**(8), 1277–1282.
- Meyer, P. D., Valocchi, A. J. and Eheart, J. W.: 1994, 'Monitoring network design to provide initial detection of groundwater contamination', *Water Resour. Res.* **30**(9), 2647–2659.
- Pardo-Igúzquiza, E.: 1998, 'Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing', *J. Hydrol.* **210**, 206–220.
- Rodríguez-Iturbe, I. and Mejía, J. M.: 1974a, 'The design of rainfall networks in time and space', *Water Resour. Res.* **10**(4), 713–728.
- Rodríguez-Iturbe, I. and Mejía, J. M.: 1974b, 'On the transformation of point rainfall to areal rainfall', *Water Resour. Res.* **10**(4), 729–736.
- Rouhani, S. and Fiering, M. B.: 1986, 'Resilience of a statistical sampling scheme', *J. Hydrol.* **89**, 1–11.
- Rouhani, S. and Hall, T.: 1988, 'Geostatistical schemes for groundwater sampling', *J. Hydrol.* **103**, 85–102.
- Shafike, N. G., Duckstein, L., and Maddock, T.: 1992, 'Multicriterion analysis of groundwater contamination management', *Water Resour. Bull.* **28**(1), 33–43.
- Shannon, C. E.: 1948, 'A mathematical theory of communication', *Bell Syst. Tech. J.* **27**, 379–423.
- Shannon, C. E. and Weaver, W.: 1949, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Illinois, U.S.A.
- Tiedman, C. and Gorelick, S.: 1993, 'Analysis of uncertainty in optimal groundwater contaminant capture design', *Water Resour. Res.* **29**(7), 2139–2153.
- Wagner, B. J. and Gorelick, S. M.: 1987, 'Optimal groundwater quality management under parameter uncertainty', *Water Resour. Res.* **23**(7), 1162–1174.
- Wagner, B. J. and Gorelick, S. M.: 1989, 'Reliable aquifer remediation in the presence of spatially variable hydraulic conductivity: From data to design', *Water Resour. Res.* **25**(10), 2211–2225.
- Woldt, W. and Bogardi, I.: 1992, 'Ground water monitoring network design using multiple criteria decision making and geostatistics', *Water Resour. Bull.* **28**(1), 45–62.
- Yfantis, E. A., Flatman, G. T. and Behar, J. V.: 1987, 'Efficiency of kriging estimation for square, triangular, and hexagonal grids', *Math. Geol.* **19**, 183–205.